中華農業研究 (Jour. Agric. Res. China) 45(4)：336～351 (1996)

區域試驗不均衡資料之穩定性子集合分析[1]

呂 秀 英[2]

摘要：本省區域試驗一般採用的穩定性介量多以最受廣泛使用的直線迴歸分析法爲主。該法僅適用於每一基因型與環境組合都有資料之場合，因此一旦含有缺區，便無法正確估得環境指標，並進行穩定性分析；過去一般處理缺區的方式，係將含缺區的品系或環境全部刪除，使其呈均衡資料後，再進行穩定性分析，如此在缺區數多時勢必會造成品系或環境數減少，反而降低了評估結果的準確度。對此，本研究利用子集合分析法以擴大穩定性分析之應用範圍至不均衡的區域試驗資料。子集合分析可處理含交感模式之不均衡資料，在觀念上相當簡單直接，係就所分割成的兩個不同子集合均衡資料各進行穩定性分析，並利用兩子集合重疊部份，對共存環境下之品系進行穩定性選拔。惟該法在使用上必須注意，所有品系的共存環境數不得太少。

關鍵詞：子集合分析、穩定性、區域試驗、不均衡資料、迴歸。

育種程序中區域試驗施行之目的，在選出高產且穩定之品種。本省區域試驗一般採用的穩定性介量多以最受廣泛使用的直線迴歸分析法爲主，即必須對每一基因型以各環境所有重複之平均性狀觀測值隨環境指標 (environmental index) 作迴歸。目前大部份的穩定性分析法僅能適用於每一基因型與環境組合都有資料之場合，尤其是直線迴歸分析模式，因此一旦含有缺區，便無法正確估得環境指標，並進行穩定性分析；至於每一基因型與環境組合都有資料，但重複不等，因爲仍然是以其重複平均值進行迴歸分析，故無影響，在此並不將之視爲區域試驗穩定性分析的缺區現象。過去一般處理缺區的方式，係將含缺區的品系或環境全部刪除，使其呈均衡資料後，再進行穩定性分析，如此在多量缺區的情況下勢必會犧牲不少辛苦調查收集之試驗資料，而且一旦欲刪除的品系爲重要品系或環境數原本即少的情形下，尤其可惜，同時也因爲品系或環境數減少，反而降低了評估結果的準確度。對此呂[2]曾考慮在不摒除任何品系及環境數之情況下，利用傳統的缺區估計法使塡入缺區估計值後，以均衡資料的型態再進行標準穩定性分析，結果發現，雖然缺區估計法相當簡單，但因爲會造成缺區品系之穩定性係數低估，故僅適用於穩定品系含少數一、二個缺區的場合，一旦缺區發生在不穩定品系時，容易使原來不穩定品系被誤認爲是穩定。這是由於以無交感模式來預估實際上交感顯著存在的區域試驗資料之缺區值，是不恰當的，故並不適用於試驗中含有對環境反應較敏感的不穩定品種[2]。至於其他以塡入估計值進行缺區矯正的程序，例如常數配合法 (fitting constant method) [9,14,18,21]，因爲也是用以分析處理組合因子間無交感效應時的複因子試驗不均衡資料，仍會使得穩定性分析的結果產生偏誤[5]。

基本上，針對部份缺區之統計方法，可廣泛地分爲以下幾種類別，而彼此間互不排斥[12,13]：(一)放棄不均衡記錄單位而僅就均衡的部份資料進行分析 (complete-case analysis) ；(二)利用有用之最

1. 臺灣省農業試驗所　研究報告第1869號。
2. 本所農藝系副研究員。臺灣省　臺中縣　霧峰鄉。

大資料集合來估計個別介量的方法（available-case methods）；(三)填入估計值使呈均衡資料後以標準方法分析的程序（imputation-based procedures）；(四)加權的程序（weighting procedures）；(五)以定義一缺區資料之模式爲基礎的程序（model-based procedures）。第一類方法最爲簡單，但會損失所拋棄之不均衡資料的相關訊息。上述所提之缺區估計屬於第三類。至於第四類則涉及較繁複的統計理論及計算程序，而且效率亦有待質疑，例如將加權值引入迴歸分析模式中，利用迭代接近法（iterative method）求得穩定性迴歸估值時，必須注意迭代接近法可能有無法收斂的問題[5]。第五類程序中以常數配合法及最大可能估計法（maximum likelihood estimation）爲最常見的方法；常數配合法係假設在一個複因子試驗無交感效應存在之約束模式（restricted model）下僅分析主效應的方法，不適用於含交感效應的情形[5]；而將缺區視爲介量以求均衡資料之最大可能函數的方法，雖然有其優點[11]，但僅在某些特殊場合有效，且一般引用於不均衡資料之分析並不普遍可靠，除非樣品大小增加到缺區所佔比例已近於 0 時，否則該法並不能獲致適當的性質[12]。因此，本研究主要針對第二類的缺區處理策略在區域試驗不均衡資料之適用性進行探討。在不摒除現有品系及環境之原則下，子集合分析（subset analysis）或許是第二類中可以考慮採用的方法之一，藉此以擴大穩定性分析之應用範圍至不均衡的基因型與環境組合資料。

統計方法

一、穩定性迴歸分析法

將基因型 i（i=1,…,m）在環境 j（j=1,…,n，爲地區與年度之組合）下的平均性狀觀測資料 y_{ij} 以下列數學模式表之：

$$y_{ij}=\mu+G_i+E_j+I_{ij}+e_{ij} \qquad [1]$$

式中

$\hat{\mu}$＝總平均值＝$\sum_{i=1}^{m}\sum_{j=1}^{n}y_{ij}/mn=\bar{y}_{..}$

$\hat{G}_i$＝基因型 i 的遺傳效應＝$\bar{y}_{i.}-\bar{y}_{..}$，而 $\bar{y}_{i.}=\sum_{j=1}^{n}y_{ij}/n$

$\hat{E}_j$＝環境 j 的效應＝$\bar{y}_{.j}-\bar{y}_{..}$，而 $\bar{y}_{.j}=\sum_{i=1}^{m}y_{ij}/m$

$\hat{I}_{ij}$＝基因型 i 與環境 j 之交感效應＝$y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..}$

e_{ij}＝試驗機差，而符合平均值爲 0，變方爲σ^2之常態分布。

並假定：

$\sum_{i=1}^{m}G_i=0$，$\sum_{j=1}^{n}E_j=0$，$\sum_{i=1}^{m}I_{ij}=\sum_{j=1}^{n}I_{ij}=0$，亦即假定 G_i，E_j 及 I_{ij} 均爲固定型效應[20]。

由於交感效應多呈環境指標的直線函數，

$$I_{ij}=\beta_iE_j+\theta_{ij} \qquad [2]$$

故直線迴歸模式可寫成

$$y_{ij}=\mu+G_i+E_j+\beta_iE_j+\theta_{ij}+e_{ij} \qquad [3]\ [15]$$

$$=\mu+G_i+(1+\beta_i)E_j+\theta_{ij}+e_{ij} \qquad [4]\ [8]$$

$$=a_i+b_iE_j+\theta'_{ij} \qquad [5]\ [6]$$

式中 θ_{ij} 爲離迴歸機差，$a_i=\bar{y}_{i.}$，b_i 爲各基因型平均性狀觀測值隨環境指標作迴歸所得的迴歸係數，一般皆以最小平方法（LS）來估計。農業育種學家多以 b_i 定義爲基因型 i 的穩定性係數（stability index）[6,8]：b=1 爲合乎一般農業要求的穩定基因型，其表現在有利環境下可獲改善，但處

於不利時也不致於太差；$0 \leq b < 1$ 爲符合生物觀點之穩定性，該類基因型之表現在優勢中也許不十分傑出，但在劣勢裡可保持相當的忍耐力；$b > 1$ 則爲不穩定或可塑的基因型，其對環境之變異相當敏感，環境之改善固然可獲致最大的增進，但環境變壞，其表現亦差。至於選取何者爲選拔準則，端視實際需要而決定。倘以 G×E 交感效應隨環境指標作迴歸所得的迴歸係數β而言[15]，則相當於以$\beta = -1$或$\beta = 0$來選拔出符合生物或農業觀點的穩定品系。

二、子集合分析

子集合分析是專門用來處理不均衡資料中模式有交感的其中一種統計方法，又稱爲格均值分析（cell means analysis）。在變方分析（analysis of variance）中，若觀測值被安排在雙向配置（two-way layout）之表格內時，其安放之位置形如格子狀，爲最小之試驗單位，則稱此一方格爲「格」。在變方分析中，每一格內之觀測值所取的平均值，謂爲格均值。複因子試驗中各格（或稱各處理組合）之觀測值數相等的資料，稱爲均衡資料；若各格之觀測值數不等，稱爲不均衡資料。不均衡資料中至少有兩類特殊情形，係屬於均衡資料之標準分析範圍內，即規劃上之不均衡性及缺區，前者例如拉丁方設計、均衡不完全區集試驗等，後者爲資料及規劃上皆形同均衡，但僅少數觀測值缺失之情況。眞正不均衡資料係摒除以上兩種不均衡性外，爲格內觀測值數不相等（含很多格內無觀測值）的情形，而又可分爲兩類：所有格塡滿的資料，以及某些空格的資料。各種類別的不均衡資料，可根據因子間交感效應存在與否，各有不同的適應分析方法：當所有格塡滿資料時，交感模式以均值加權平方法（weighted squares of mean analysis）或格均值分析進行，而無交感模式則只行主效應分析；若是某些爲空格的資料且模式有交感時，僅能使用格均值分析[17]。

區域試驗實際上是一個含有品種及地區兩因子的複因子試驗，其線性模式中並含有基因型與環境之交感效應，因此觀測值 Y 在 m－n 單位（假設有 m 品種 n 環境）上的缺區現象，應該視爲含交感之某些空格的不均衡資料，故可以採用格均值分析中的子集合分析。

所謂子集合分析，就是將 m－n 單位（m×n 兩向表）的不均衡資料集分割成均衡的資料集，這兩個資料集各爲原來資料集的一個子集合，然後再就各子集合分別進行統計分析與結果解釋。事實上，這些子集合資料的分析因互有重疊，故並不互相獨立，但分析上顯然變得極爲簡單，而且個別之解釋上也很直接。

材料及方法

本研究利用呂及曾[1]之飼料玉米綜合新品系區域試驗資料爲分析材料，以檢視本研究所擬方法的適用性。試驗於1986年春作至1987年秋作進行，計二年四期作。參試品系包括 PC8、PC7、PCC、PUC 及對照品系（CK）之臺南選10號，分別於臺東縣東河鄉、臺東縣鹿野鄉及花蓮縣萬榮鄉共 3 個地區進行試驗。新品系 PC8 及 PC7 均爲自國際玉米小麥研究中心（International Maize and Wheat Improvement Center, CIMMYT）引進之抗倒伏的熱帶型半硬粒種，PCC 及 PUC 則爲本省朴子分場培育之熱帶型硬粒種。各試驗之田間設計採用逢機完全區集法（RCBD），四重複，按當地高級產量試驗標準，六行區，行長 5m，行株距 75×25cm，小區面積 22.5m^2，田間作業管理採一般方式執行，調查公頃籽實產量（kg/ha）。

區域試驗所調查的產量結果及環境指標估值列於表 1。本研究之玉米區域試驗係以地區及年度之組合作爲環境變異，故合計 5 個品系 6 個環境進行穩定性分析。爲方便說明，並使一目瞭然，以下相關圖表及其說明將簡化符號，設品系代號各爲 V1～V5，而環境代號各爲 E1～E6。不均衡資料組則是利用人爲的方式，刪去該 5 × 6 兩向表資料中的一個或一個以上的單位（格）資料，使造成缺區現象。爲探討缺區發生在穩定或不穩定品種時，其對各種理論分析所造成的可能影響，因此根據均衡資料的穩定性分析結果，分別就表現出不同穩定度（穩定性係數 $b < 1$、$b = 1$、$b > 1$）的品種，逢機

刪去部份資料，以模擬成各種含有缺區的不均衡資料組。這些不均衡資料組均將利用子集合分析，先將其分割成兩個均衡的子集合後，再就各子集合分別進行標準的穩定性迴歸分析，利用兩子集合資料互有重疊之部份，對所有參試品系在共存環境下的綜合表現作一整合性的推論，以選出理想的優良品種。

表1. 春作五個玉米品系於六個環境（兩年×三地區）下生長的四重複平均籽實產量資料（公斤／公頃）

Table 1. Mean kernel yield (kg/ha) of four replicates on five maize varieties grown in six environments (2 years×3 sites) under spring crop season

Variety	Environment[y]					
	E1	E2	E3	E4	E5	E6
V1 (PC8)	3,289	2,163	3,692	3,921	1,589	1,805
V2 (PC7)	3,487	1,660	3,148	3,897	1,937	1,404
V3 (PCC)	2,651	1,996	2,375	2,920	982	1,158
V4 (PUC)	2,467	2,095	2,575	3,743	1,846	1,489
V5 (CK)[z]	2,753	1,761	3,913	4,179	1,890	1,678
Average	2,929	1,935	3,141	3,732	1,649	1,507
Envir. index	447	−547	659	1,250	−833	−975

[z] CK：臺南選10號 (Tainan sel. No. 10)
[y] E1：1986臺東縣東河鄉 (Tongho, Taitung)
E2：1986臺東縣鹿野鄉 (Louyai, Taitung)
E3：1986花蓮縣萬榮鄉 (Wanlong, Hualien)
E4：1987臺東縣東河鄉 (Tongho, Taitung)
E5：1987臺東縣鹿野鄉 (Louyai, Taitung)
E6：1987花蓮縣萬榮鄉 (Wanlong, Hualien)

結　果

一、原始均衡資料的穩定性分析結果

本研究飼料玉米綜合新品系區域試驗以均衡資料進行之綜合變方分析結果列於表 2，品系與環境之交感效應呈極顯著，顯示參試品系間對環境的反應不一致，其穩定性分析的結果，如表 3 所示。由表 3 得知，各品系的分析結果均滿足直線迴歸法應用於穩定性分析的兩個重要統計前提，即各品系之迴歸係數 (b) 均顯著存在（直線關係成立），同時所有品系之迴歸離差均方 (S^2d) 不顯著且互爲均質，因此所求得之迴歸係數估值 (b) 足以解釋基因型對環境之適應能力。各參試品系之 b 值間的差異未達0.05之顯著水準，但如需進一步再互作比較，四個新品系的 b 值均相對地較臺南選10號之對照品系 (CK) 爲低，即表現可更爲穩定，其中 PC8 及 PC7 之 b 值最接近於1，爲符合農業觀念之穩定品種，其會隨著環境改善而表現增進，但也不致因爲環境不利而表現太差，PCC 及 PUC 之 b 值爲0.81，小於1，對環境的變異亦相當穩定，尤其在惡劣環境下可較 PC8 及 PC7 穩定，但因環境改善所能提高的產量卻不如 PC8 或 PC7 高，爲符合生物觀念的穩定性；對照品系 (CK) 的 b 值大於1，表示其穩定性相對地較其他四個新育成品系爲低。由於育種上欲選拔優良的品系，須兼顧穩定性及生產力，故一般均同時以穩定性介量與產量平均值爲選拔依據，如表 3 所示，如以優於對照品系爲選拔之目標，而 PC8 應爲最理想的優良品系，其有較高的產量及穩定性。

表2. 玉米品系春作籽實產量區域試驗之綜合變方分析結果

Table 2. Combined analysis of variance for kernel yield of maize varieties in a regional trial under spring crop season

Source	D. F.	Mean squares
Block/Environment	18	125,359.5
Environment (E)	5	16,057,406.8**
Variety (V)	4	2,241,246.5**
V×E	20	458,425.3**
Error	71	174,205.0

** significant at 1% level.

表3. 玉米品系春作籽實產量區域試驗以直線迴歸法進行穩定性分析之結果

Table 3. Stability analysis of kernel yield of maize varieties in a regional trial under spring crop season using linear regression method

Variety	Mean yield (a)	Stability index (b)	Deviation from regression (S^2d)
V1 (PC8)	2,743 a[z]	1.10**	41,876
V2 (PC7)	2,589 ab	1.11**	96,429
V3 (PCC)	2,014 c	0.81**	107,299
V4 (PUC)	2,369 b	0.81**	92,001
V5 (CK)	2,695 a	1.17**	136,976
Average	2,482	1.00	94,917
LSD (0.05)	237	0.31	
Bartlett's χ^2			1.25[ns]

[z] Means with the same letter of a column are not significantly different at 5% level by LSD Test.
** significant at 1% level under hypothesis b=0.
[ns] nonsignificant at 5% level.

二、不均衡資料的子集合穩定性分析

(一)單一缺區之場合

首先考慮只有一個缺區的最簡單場合，即只有某一環境缺少一個品系。缺區發生處，依上述玉米區域試驗均衡資料之穩定性分析結果，假設其可能是發生在符合生物穩定（b<1）、農業穩定（b=1）或不穩定（b>1）品種的某個環境上。選定這三種不同穩定度的玉米參試品系，分別爲 V1（PC8）、V3（PCC）及 V5（CK）。至於是在哪一個環境缺失資料，則於六個環境中逢機決定。因此所模擬成的單一缺區之三種規劃情形，如圖 1 所示，圖中格內✓符號表示有資料，空白則代表無資料。利用子集合分析，將原來 5 × 6 兩向表之不均衡資料集分割成兩個 5 × 5 及 4 × 6 的子集合，其中第一個 5 × 5 子集合爲包含所有品系但刪去含缺區環境的均衡資料集，而第二個 4 × 6 子集合爲包含所有環境但刪去含缺區品系的均衡資料集。然後再就各子集合分別進行穩定性分析，其所估得的環境指標、產量平均值、穩定性係數及迴歸離差均方的估計值，分別列於表 4 及表 5。

(A)

5 x 6

	V1	V2	V3	V4	V5
E1	v	v	v	v	v
E2	v	v	v	v	v
E3	v	v	v	v	v
E4	v	v	v	v	v
E5		v	v	v	v
E6	v	v	v	v	v

→

5 x 5

	V1	V2	V3	V4	V5
E1	v	v	v	v	v
E2	v	v	v	v	v
E3	v	v	v	v	v
E4	v	v	v	v	v
E6	v	v	v	v	v

4 x 6

	V2	V3	V4	V5
E1	v	v	v	v
E2	v	v	v	v
E3	v	v	v	v
E4	v	v	v	v
E5	v	v	v	v
E6	v	v	v	v

(B)

5 x 6

	V1	V2	V3	V4	V5
E1	v	v	v	v	v
E2	v	v	v	v	v
E3	v	v	v	v	v
E4	v	v	v	v	v
E5	v	v		v	v
E6	v	v	v	v	v

→

5 x 5

	V1	V2	V3	V4	V5
E1	v	v	v	v	v
E2	v	v	v	v	v
E3	v	v	v	v	v
E4	v	v	v	v	v
E6	v	v	v	v	v

4 x 6

	V1	V2	V4	V5
E1	v	v	v	v
E2	v	v	v	v
E3	v	v	v	v
E4	v	v	v	v
E5	v	v	v	v
E6	v	v	v	v

(C)

5 x 6

	V1	V2	V3	V4	V5
E1	v	v	v	v	v
E2	v	v	v	v	v
E3	v	v	v	v	v
E4	v	v	v	v	
E5	v	v	v	v	v
E6	v	v	v	v	v

→

5 x 5

	V1	V2	V3	V4	V5
E1	v	v	v	v	v
E2	v	v	v	v	v
E3	v	v	v	v	v
E5	v	v	v	v	v
E6	v	v	v	v	v

4 x 6

	V1	V2	V3	V4
E1	v	v	v	v
E2	v	v	v	v
E3	v	v	v	v
E4	v	v	v	v
E5	v	v	v	v
E6	v	v	v	v

圖1. 玉米區域試驗不同品系含單一缺區時的不均衡資料之子集合分析
(缺區各為A：y_{15}, B：y_{35}, C：y_{54})

Fig. 1. Subset analysis of unbalanced data with only one missing plot occurred on different varieties in a maize regional trial.
(Missing plot occurred on A:y_{15}, B:y_{35}, C:y_{54})

表4. 玉米區域試驗單一缺區之不均衡資料各子集合的環境指標估值

Table 4. Estimates of environmental index for individual subset of unbalanced data with only one missing plot in a maize regional trial

Envir.	V1[z]		V3		V5	
	I[y] 5×5	II 4×6	I 5×5	II 4×6	I 5×5	II 4×6
E1	281	423	281	400	697	545
E2	−714	−539	−714	−679	−297	−450
E3	492	586	492	733	909	519
E4	1,083	1,268	1,083	1,336		1,192
E5		−753		−784	−583	−840
E6	−1,142	−985	−1,142	−1,005	−725	−965
Max.	1,083	1,268	1,083	1,336	909	1,192
Min.	−1,142	−983	−1,142	−1,005	−725	−965
Range	2,225	2,253	2,225	2,341	1,634	2,157

[z] genotype with missing plot.
[y] subset of data.

表5. 玉米區域試驗單一缺區之不均衡資料各子集合的產量平均值 、迴歸離差均方及穩定性係數之估值

Table 5. Estimates of mean yield, deviation from regression and stability indices for individual subset of unbalanced data with only one missing plot in a maize regional trial

Variety	V1[z]		V3		V5	
	I[y] 5×5	II 4×6	I 5×5	II 4×6	I 5×5	II 4×6
	Mean yield (a)					
V1 (PC8)	2,974 a		2,974 a	2,743 a	2,508 a	2,743 a
V2 (PC7)	2,719 b	2,589 a	2,719 b	2,589 ab	2,327 ab	2,589 ab
V3 (PCC)	2,220 d	2,014 c	2,220 d		1,833 c	2,014 c
V4 (PUC)	2,474 c	2,369 b	2,474 c	2,369 b	2,094 b	2,369 b
V5 (CK)	2,857 ab	2,695 ab	2,857 ab	2,695 a	2,399 a	
Average	2,649	2,417	2,649	2,599	2,232	2,429
	Deviation from regression (S^2d)					
V1 (PC8)	26,631		26,631	59,757	29,060	58,923
V2 (PC7)	90,369	87,436	90,369	93,519	123,418	88,016
V3 (PCC)	79,876	117,530	79,876		134,581	70,599
V4 (PUC)	111,535	72,593	111,535	93,023	24,985	92,702
V5 (CK)	167,920	147,003	167,920	88,946	182,410	
Average	95,266	106,141	95,266	83,766	98,891	77,560
Bartlett's χ^2	2.83 [ns]	0.54 [ns]	2.83 [ns]	0.23 [ns]	5.30 [ns]	0.24 [ns]
	Stability index (b)					
V1 (PC8)	1.02		1.02	1.04	1.22	1.13
V2 (PC7)	1.20	1.14	1.20	1.06	1.17	1.16
V3 (PCC)	0.70	0.82		0.70	0.88	0.87
V4 (PUC)	0.85	0.84	0.85	0.77	0.56	0.84
V5 (CK)	1.22	1.20	1.22	1.14	1.16	

[z] genotype with missing plot.
[y] subset of data.
[ns] nonsignificant at 5% level.

由於穩定性係一種相對比較的概念，因此可就該兩個不同子集合的穩定性分析結果，予以個別解釋但綜合討論。換言之，吾人可以由第一個 5 × 5 子集合之分析結果，檢視出所有五個品系在其中五個環境下的穩定性表現，而由第二個 4 × 6 子集合之分析結果，比較其中四個品系在所有六個環境下的穩定性表現，由於這些子集合資料的分析互有重疊，即互不獨立，因此進一步可以利用重疊部份，對所有品系在共存環境下的綜合表現作一整合性的推論，以選出理想的優良品種。

由表 4 得知，不同子集合的環境指標估值，除了 V5 於環境 E4 有缺區之場合其兩個子集合的環境變異互有差異以外，其他兩子集合間的環境變異範圍皆頗爲接近。由表 5 得知，無論缺區發生何處，各子集合的產量平均值皆表現一致，產量由高至低，依次爲 V1、V5、V2、V4、V3。各子集合內所有品系之間的迴歸離差均方，皆呈均質，表示符合穩定性直線迴歸模式的統計前提，故所求得之穩定性介量的估計值應足以代表基因型對環境之適應能力（表 5）。由表 5 之穩定性係數估值得知，穩定的 V1（PC8）品系在環境 E5 發生單一缺區時，從子集合 I 的穩定性分析結果發現，在不計環境 E5 的情形下，V1 品系確爲一符合農業穩定的品種（b＝1），V3 及 V4 爲符合生物穩定性的品系（b<1），V2 及 V5 爲不穩定品系（b>1）；而從子集合 II 的穩定性分析結果，V3、V4 及V5 在所有六個環境下的穩定性表現與子集合 I 所獲結果一致，但 V2 的 b 值在加入環境 E5 後而有變小的趨勢，此將使得 V2 變得較爲穩定；綜合兩子集合之分析結果，可推知在不考慮環境 E5 下，所有參試品系中仍應以 V1 爲最符合農業穩定的品系，其次可考慮 V2，至於 V3 及 V4 爲符合生物穩定的品種，V5 則爲最不穩定品種。倘若缺區是發生在符合生物穩定的 V3（PCC）品系時，就兩子集合之穩定性分析分開個別解釋並綜合討論的結果，亦與前述缺區品系爲 V1 之情形所獲品種穩定性評價的結果完全相同。一旦缺區是發生在不穩定的品系（V5），從子集合 I 的穩定性分析結果發現，在不計環境 E4 的情形下，V1 之穩定性最差，在加入環境 E4 後，從子集合 II 的穩定性分析結果來看，V1 的穩定度可獲提高，至於 V2、V3 及 V4 的穩定性則維持不變；綜合兩子集合穩定性分析結果，倘不計環境 E4 下，應選擇 V3 爲最符合農業穩定的品種，因其 b 值最接近於 1，若僅就其中四個品系於所有六個環境之相對性比較，以 V1 及 V3 的穩定性較佳，但此時因缺乏了對照品系（V5），致使育種上優良品種選拔之意義降低。

按品系在重疊部份結果上的差異，則是因爲兩子集合之間的環境變異有較大的不同所致，例如前述之 V5 在環境 E4 有缺區的情形下，V1 在兩子集合間的穩定性分析結果之所以有差異（表 5），可從表 4 之各子集合的環境指標值得到解釋及說明；由表 4 得知，是否加入環境 E4，對環境變異的影響很大，致使子集合 I 的環境變異範圍甚小於子集合 II，又因爲環境 E4 之環境指標估值最大（表 1），爲一最高產環境，對於像 V1 這種實際上符合農業穩定性的品種而言，由於其會隨著環境改善而表現增進，故對高產環境的敏感度更甚於其他品系，因此所受到的影響也較大。至於品系在其他環境有缺區的情形，例如表 5 中之 V1 或 V3 於環境 E5 有缺區之場合，由於 E5 之環境指標爲負值（表 1），即爲一低產環境，但並非爲最低產（並非最大負值），加入與否對環境變異範圍影響不大，故各品系在重疊部份的結果大致相同。

㈡多個缺區之場合

進一步探討子集合分析應用於區域試驗資料含多個缺區的適用性。兩個以上的缺區發生處有三種可能：⑴至少兩個以上的環境缺少同一品系，⑵至少兩個以上的環境缺少不同的品系，⑶某一環境缺少兩個以上的品系。首先以兩個缺區的情形爲例說明，如圖 2 所示：情形 A 爲 E4及

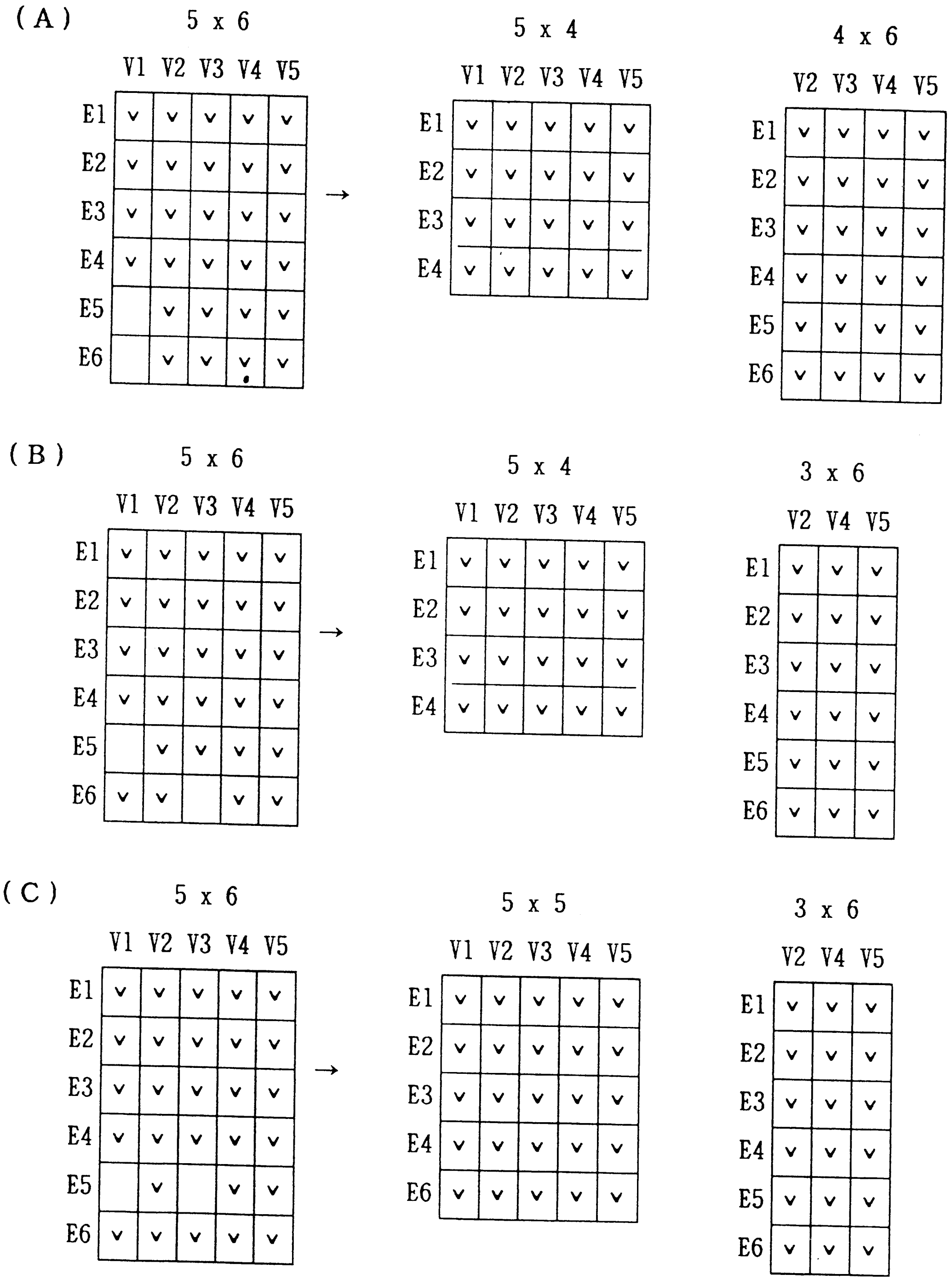

圖2. 玉米區域試驗含兩個缺區時的不均衡資料之子集合分析
(缺區各爲 A:y_{15}, y_{16}; B:y_{15}, y_{36}; C:y_{15}, y_{35})

Fig. 2. Subset analysis of unbalanced data with two missing plots in a maize regional trial.
(Missing plot occurred on A:y_{15}, y_{16} B:y_{15}, y_{36}, C:y_{15}, y_{35})

E5 環境均缺少 V1 品系的資料，利用子集合分析可將此不均衡資料分割成兩個 5 × 4 及 4 × 6 的子集合均衡資料集；情形 B 爲 E5 環境缺少 V1 品系且 E6 環境缺少 V3 品系的不均衡資料，子集合分析結果爲兩個 5 × 4 及 3 × 6 的兩個子集合均衡資料集；情形 C 爲 V1 及 V3 品系同時於 E5 環境缺少資料，利用子集合分析將其分割成兩個 5 × 5 及 3 × 6 的子集合均衡資料集，並針對這些子集合個別分開進行穩定性分析。不同子集合的環境指標估值如表 6 所示，其中以含兩個缺區環境的場合（情形 A，B）之兩子集合的環境變異間的差異較大。由表 7 得知，無論缺區發生何處，各子集合的產量平均值均表現一致，產量由高至低，依次爲 V1、V5、V2、V4、V3。各子集合內所有品系之間的迴歸離差均方，皆呈均質，表示符合穩定性直線迴歸模式的統計前提（表 7）。在情形 A 下，倘不計入 E5 及 E6 環境，以 V1 品系爲最符合農業穩定的品種，若於所有六個環境進行選拔，則僅能就 V2、V3、V4、V5 四個品系進行選拔，結果得知 V2、V3、V4 品系之穩定性都略優於 V5，而其中以 V2 最好，其 b 值最接近於 1；在情形 B 下，不計入 E5 及 E6 環境時仍以 V1 品系爲最符合農業穩定的品種，若考慮所有六個環境，則僅能就 V2、V4、V5 品系進行選拔，其中以 V2 最好；在情形 C 下，其結果與其他情形相似，即不考慮缺區環境 E5 時，應選擇 V1品系，如考慮所有環境，則以 V2 品系較其他品系爲優（表 7）。

表6. 玉米區域試驗含兩個缺區之不均衡資料各子集合的環境指標估值

Table 6. Estimates of environmental indices for individual subset of unbalanced data with two missing plots in a maize regional trial

Envir.	Case-A[z]		Case-B		Case-C	
	I[y] 5×4	II 4×6	I 5×4	II 3×6	I 5×5	II 3×6
E1	−5	423	−5	351	281	351
E2	−999	−539	−999	−712	−714	−712
E3	206	586	206	661	492	661
E4	798	1,268	798	1,389	1,083	1,389
E5		−753		−661		−661
E6		−985		−1,028	−1,142	−1,028
Max.	798	1,268	798	1,389	1,083	1,389
Min.	−999	−985	−999	−1,028	−1,142	−1,028
Range	1,797	2,253	1,797	2,417	2,225	2,417

[z] Case-A = missing plots are y_{15}, y_{16}.
Case-B = missing plots are y_{15}, y_{36}.
Case-C = missing plots are y_{15}, y_{35}.
[y] subset of data.

表7. 玉米區域試驗兩個缺區之不均衡資料各子集合的產量平均值 、迴歸離差均方及穩定性係數之估值

Table 7. Estimates of mean yield, deviation from regression and stability indices for individual subset of unbalanced data with two missing plots in a maize regional trial

Variety	Case-A^{z}		Case-B		Case-C	
	I^{y} 5×5	II 4×6	I 5×5	II 4×6	I 5×5	II 4×6
	Mean yield (a)					
V1 (PC8)	3,266 a		3,266 a		2,974 a	
V2 (PC7)	3,048 a	2,589 ab	3,048 a	2,589 ab	2,719 b	2,589 ab
V3 (PCC)	2,485 b	2,014 c	2,485 b		2,220 d	
V4 (PUC)	2,720 b	2,369 b	2,720 b	2,369 b	2,474 c	2,369 b
V5 (CK)	3,151 a	2,695 a	3,151 a	2,695 a	2,857 ab	2,695 a
Average	2,394	2,417	2,934	2,551	2,649	2,551
	Deviation from regression (S^2d)					
V1 (PC8)	39,945		39,945		26,631	
V2 (PC7)	132,186	87,436	132,186	91,139	90,369	91,139
V3 (PCC)	37,538	117,530	37,538		79,876	
V4 (PUC)	167,119	72,593	167,119	74,311	111,535	74,311
V5 (CK)	192,411	147,003	192,411	96,118	167,920	96,118
Average	113,840	106,141	113,840	87,189	95,266	87,189
Bartlett's x^2	3.99 ^{n}	0.54 ^{ns}	3.99 ^{ns}	0.07 ^{ns}	2.83 ^{ns}	0.07 ^{ns}
	Stability index (b)					
V1 (PC8)	1.02		1.02		1.02	
V2 (PC7)	1.24	1.14	1.24	1.07	1.20	1.07
V3 (PCC)	0.48	0.82	0.48		0.70	
V4 (PUC)	0.84	0.84	0.84	0.79	0.85	0.79
V5 (CK)	1.41	1.20	1.41	1.14	1.22	1.14

^{z} Case-A = missing plots are y_{15}, y_{16}.
Case-B = missing plots are y_{15}, y_{36}.
Case-C = missing plots are y_{15}, y_{35}.
^{y} subset of data.
^{ns} nonsignificant at 5% level.

繼之，再探討子集合分析應用於含三個缺區的區域試驗資料之適用性，仿照上述所提之兩個以上缺區發生的三種可能，列舉如圖 3 所示之三個缺區的五種可能發生情形。這五種缺區發生情形，利用子集合分析可將原來不均衡資料分別分割成兩個子集合之均衡資料，然後再就各子集合個別進行穩定性分析。子集合所含環境數愈少，環境變異範圍愈小，表示這些少數參試環境難以適當地代表整個環境族群（表 8 ）。在考慮所有品系間的產量平均值比較時，仍以 V1 爲最高產的品系（表 9 ）。各子集合均衡資料進行穩定性分析之結果，情形 A、B、D 下在子集合 I 的品系間迴歸離差均方呈異質，即無法符合穩定性直線迴歸模式的統計前提，這是因爲該等子集合所含的環境數太少所致，因此其求得之穩定性係數估值不足以代表基因型對環境之適應能力；換言之，僅有 C、E 之缺區情形，能夠利用子集合分析獲得有效的穩定性評估結果（表 9 ）。在情形 C、E 下，倘不計入缺區環境，由子集合 I 比較所有品系的穩定性表現，可看出應選擇 V1 品系爲最理想的品種，其 b 值最接近於 1 ；但如欲考慮所有環境，則情形 C 只能就 V2、V4、V5 三個品系進行比較，而其中以 V2 較符合農業穩定，但其平均產量低於 V5 之對照品系；至於情形 E 下，倘欲考慮所有環境，則只能就 V4 及 V5 (CK) 進行比較（表 9 ）。

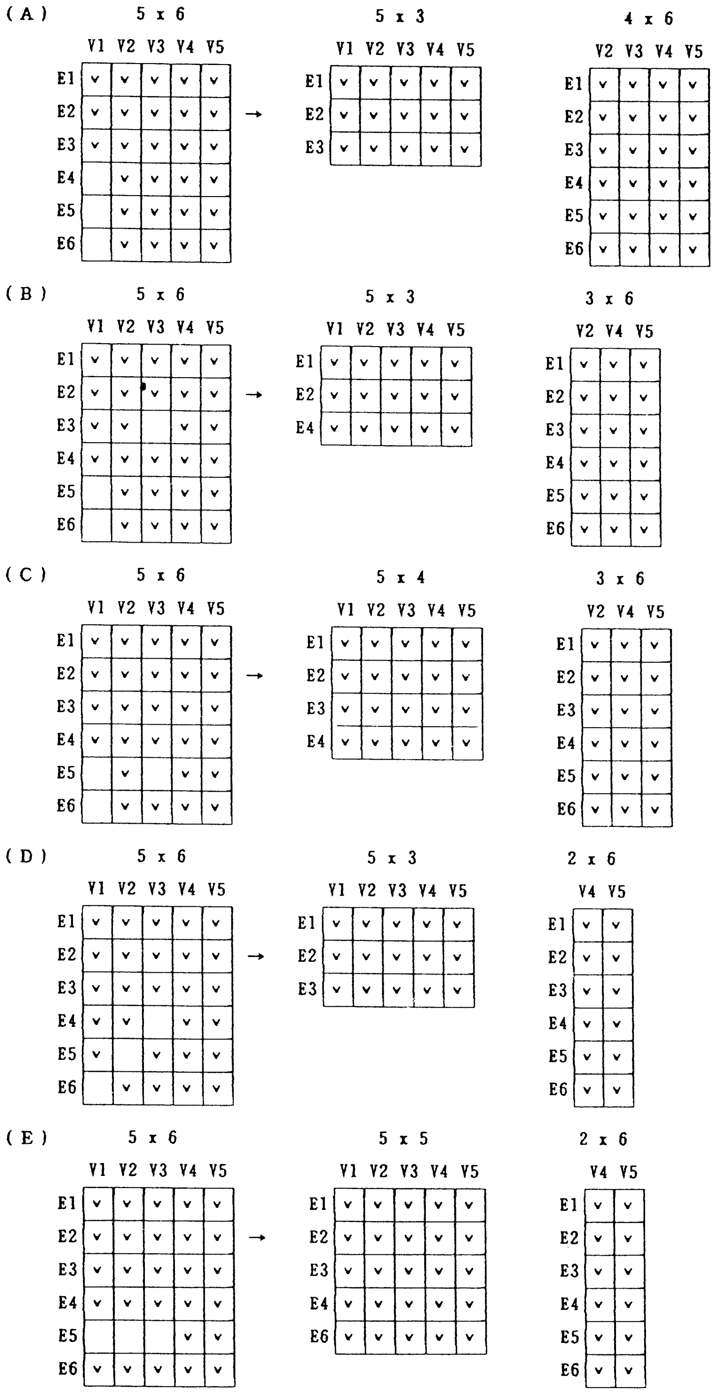

圖3. 玉米區域試驗含三個缺區時的不均衡資料之子集合分析
(缺區各爲 A:y_{14}, y_{15}, y_{16} ; B:y_{15}, y_{16}, y_{33} ; C:y_{15}, y_{16}, y_{35} ; D:y_{16}, y_{25}, y_{34} ; E:y_{15}, y_{25}, y_{35})

Fig. 3. Subset analysis of unbalanced data with three missing plots in a maize regional trial. (Missing plots occurred on A:y_{14}, y_{15}, y_{16} ; B:y_{15}, y_{16}, y_{33} ; C:y_{15}, y_{16}, y_{35} ; D:y_{16}, y_{25}, y_{34} ; E:y_{15}, y_{25}, y_{35})

表8. 玉米區域試驗含三個缺區之不均衡資料各子集合的環境指標估值

Table 8. Estimates of environmental indices for individual subset of unbalanced data with three missing plots in a maize regional trial

Envir.	Case-A^z		Case-B		Case-C		Case-D		Case-E	
	I^y 5×3	II 4×6	I 5×3	II 3×6	I 5×4	II 3×6	I 5×3	II 2×6	I 5×5	II 2×6
E1	261	423	64	351	−5	351	261	78	281	78
E2	−733	−539	−930	−712	−999	−712	−733	−604	−714	−604
E3	472	586		661	206	661	472	711	492	711
E4		1,268	867	1,389	798	1,389		1,429	1,083	1,429
E5		−753		−661		−661		−665		−665
E6		−985		−1,028		−1,028		−949	−1,142	−949
Max.	472	1,268	867	1,389	798	1,389	472	1,429	1,083	1,429
Min.	−733	−985	−930	−1,028	−999	−1,028	−733	−949	−1,142	−949
Range	1,205	2,253	1,797	2,417	1,797	2,417	1,205	2,378	2,225	2,378

^{z} Case-A = missing plots are y_{14}, y_{15}, y_{16}.
Case-B = missing plots are y_{15}, y_{16}, y_{33}.
Case-C = missing plots are y_{15}, y_{16}, y_{35}.
Case-D = missing plots are y_{16}, y_{25}, y_{34}.
Case-E = missing plots are y_{15}, y_{25}, y_{35}.
^{y} subset of data.

表9. 玉米區域試驗含三個缺區之不均衡資料各子集合的產量平均值、迴歸離差均方及穩定性係數之估值

Table 9. Estimates of mean yield, deviation from regression and stability indices for individual subset of unbalanced data with three missing plots in a maize regional trial

Variety	Case-A^z		Case-B		Case-C		Case-D		Case-E	
	I^y 5×3	II 4×6	I 5×3	II 3×6	I 5×4	II 3×6	I 5×3	II 2×6	I 5×5	II 2×6
	Mean yield (a)									
V1 (PC8)	3,048 a		3,124 a		3,266 a		3,048 a		2,974 a	
V2 (PC7)	2,765 b	2,589 ab	3,015 ab	2,589 ab	3,048 a	2,589 ab	2,765 b		2,719 b	
V3 (PCC)	2,341 c	2,014 c	2,522 c		2,485 b		2,341 b		2,220 d	
V4 (PUC)	2,379 c	2,369 b	2,768 bc	2,369 b	2,720 b	2,369 b	2,379 c	2,369 b	2,474 c	2,369 b
V5 (CK)	2,809 ab	2,695 a	2,898 ab	2,695 a	3,151 a	2,695 a	2,809 ab	2,695 a	2,857 ab	2,695 a
Average	2,668	2,417	2,865	2,551	2,934	2,551	2,668	2,532	2,649	2,532
	Deviation from regression (S^2d)									
V1 (PC8)	10,798		15,601		39,945		10,798		26,631	
V2 (PC7)	210,264	87,436	230,733	91,139	132,186	91,139	210,264		90,369	
V3 (PCC)	68,765	117,530	13,762		37,538		68,765		79,876	
V4 (PUC)	324	72,593	193,725	74,311	167,119	74,311	324	61,190	111,535	61,190
V5 (CK)	358,635	147,003	79,762	96,118	192,411	96,118	358,635	61,190	167,920	61,190
Average	129,757	106,141	106,717	87,189	113,840	87,189	129,757	61,190	95,266	61,190
Bartlett's x^2	27.69 **	0.54^{ns}	10.53 *	0.07^{ns}	3.99^{ns}	0.07^{ns}	27.69 **	0^{ns}	2.83^{ns}	0^{ns}
	Stability index (b)									
V1 (PC8)	1.23		0.98		1.02		1.23		1.02	
V2 (PC7)	1.42	1.14	1.27	1.07	1.24	1.07	1.42		1.20	
V3 (PCC)	0.42	0.82	0.52		0.48		0.42		0.70	
V4 (PUC)	0.39	0.84	0.90	0.79	0.84	0.79	0.39	0.81	0.85	0.81
V5 (CK)	1.54	1.20	1.33	1.14	1.41	1.14	1.54	1.19	1.22	1.19

^{z} Case-A = missing plots are y_{14}, y_{15}, y_{16}.
Case-B = missing plots are y_{15}, y_{16}, y_{33}.
Case-C = missing plots are y_{15}, y_{16}, y_{35}.
Case-D = missing plots are y_{16}, y_{25}, y_{34}.
Case-E = missing plots are y_{15}, y_{25}, y_{35}.
^{y} subset of data.
*,** significant at 5% and 1% level, respectively.
^{ns} nonsignificant at 5% level.

當然，子集合分析也能推至三個以上缺區的場合，但一旦子集合的環境數過少，則進行穩定性迴歸分析的準確度會降低，所得分析結果便不足為信，因此建議利用子集合分析以處理不均衡區域試驗資料時，在使用上應注意分割後的子集合所含之品系及環境數不得太少，至少要在 3 個以上，才能進行子集合分析。當然，倘若分割後進行穩定性分析的品系及環境數越多，所評估之品系穩定性的結果也越是準確，也才值得信賴。

討　　論

綜合上述結果得知，由於子集合分析是使用標準的均衡資料進行統計分析，毋須修正資料及模式即可運用，故在觀念上最為簡單直接，但必須就所分割成的兩個不同子集合之穩定性分析結果，予以個別解釋，不過由於這些子集合資料的分析互有重疊，故可以進而利用重疊部份，對所有品系在共存環境下的綜合表現作一整合性的推論，以選出理想的優良品種。為避免因品系及環境數太少而影響穩定性分析之準確度，在使用上應注意子集合所含的品系及環境數，至少要在 3 個以上。另外，利用子集合分析，在拋棄不均衡資料的同時，則有損失訊息的潛在不利點，特別是在含缺區之品系或環境數很多時，樣本大小之損失將相當可觀。某些空格資料以子集合分析進行分析，雖不如均衡資料之變方分析或不均衡但格內均填滿資料之均值加權平方法可獲致較多的訊息，但效率更高，也較容易了解及解釋，同時也可利用子集合分析來檢視交感效應[17]。有些時候，不均衡資料必須經由兩次以上的分割，始能產生有效的子集合。然而，子集合分析仍可能存在以下困難[17]：(1)並非所有的資料都能被分割成有效的子集合，此時唯一適當的處理方式只能以無交感模式的型態來執行；(2)資料可能被分割成兩種以上完全不同的子集合分析，因此一般而言，某些空格之不均衡資料很少有唯一而確實的統計分析，兩位統計學家可能會對同一套資料有不同方式的分析建議，但並無絕對之對錯，重點端視研究者的興趣在何種比較而已。

導致缺區資料的機制已知或未知，是選擇適當的分析方法及結果解釋的主要因素。由於區域試驗中發生缺區，可能是由於病蟲害、鳥害、機械傷害等因素以致無法收穫，或有記錄但遺失了，或因為離值被拋棄不用，故其機制在統計上多屬於逢機性型態（missing at random, MAR）[2]，即觀測值 y_{ij} 之缺區可能發生在任一個自變數上，y_{ij} 缺失的機率與，y_{ij} 值無關。在此情況下，可以利用縮減資料之方式使資料呈均衡或近似均衡的型態來分析及敘述，以簡化計算[12]。因此，將子集合分析引用於區域試驗不均衡資料，是可以符合統計理論之要求。反之，倘導致缺區的機制不是樣本選擇之過程，即並非逢機缺失時，該缺區是不可忽略的，此時縮減資料分析是不允許的，因為會導致估計值偏誤。當區域試驗資料不符合 MAR 時，倘缺區之發生僅與 X 之自變數（環境）有關，傳統的條件推定程序[12,13]只在環境指標能以實際的外在氣候因子來取代時，才能有效地藉由外在環境變數下來推定缺區，此固然有很多優點[3,7,10,16]，但在很多場合，環境的生產水準不可能單純地只由一些有限的環境因子所決定[4]；倘缺區之發生同時與 Y 及 X 有關，即缺區機制為不可忽略，則為最難分析的一種，其涉及的問題相當複雜[12,19]。

綜言之，無論是採用何種策略來處理區域試驗缺區的問題，僅利用少數的環境是不可能計算出有效的穩定性估計值。再者，由於迴歸穩定性係數只是提供參試基因型之間的相對等級比較，所以並非一種絕對性的測量值，其所評估的穩定性結果也僅限於該試驗下的環境及品種系統；如果增加極端的環境或品種系統，則穩定性指標會因環境平均值或環境指標平方和之增加（後者會導致迴歸係數之標準機差減少）而改變估計值。此亦說明了利用子集合分析，倘子集合含有極端環境（如本研究之最高產環境），對品系之穩定性評估結果會有差異性的影響。當然，包含的極端環境愈多，交感效應呈非直線的可能性亦增加，此也將使得直線迴歸模式應用於穩定性分析的有效性降低，以致無法解釋基因型對環境之適應能力。

引用文獻

1. 呂秀英、曾清田。1993。非介量穩定性分析之研究 II. 農業觀念之穩定性選拔。中華農學會報 164：1—16。
2. 呂秀英。1996。作物區域試驗穩定性分析之缺區估計。中華農業研究 45：218—229。
3. Becker, H. C. and J. Léon. 1988. Stability analysis in plant breeding. Plant Breed. 101：1—23.
4. Beckett, J. L. 1982. Variety×environment interactions in sugar beet variety trials. J. Agric. Sci., Camb. 98：425—435.
5. Digby, P. G. 1979. Modified joint regression analysis for incomplete variety×enviornment data. J. Agric. Sci. Camb. 93：81—86.
6. Eberhart, S. A. and W. A. Russell. 1966. Stability parameters for comparing varieties. Crop. Sci. 6：36—40.
7. Eberhart, S. A., L. H. Penny and M. N. Harrison. 1973. Genotype by environment interactions in maize in Eastern Africa. East Afric. Agric. For. J. 39：61—71.
8. Finlay, K. W. and G. N. Wilkinson. 1963. The analysis of adaptation in a plant breeding programme. Austral. J. Agric. Res. 14：742—754.
9. Freeman, G. H. 1975. Analysis of interactions in incomplete two-way tables. Appl. Statis. 24：46—55.
10. Hadjichristodoulou, A. and A. Kari. 1984. Karpasia, a high yielding new durum wheat variety with improved grain quality characteristics. Techn. Bull. Agric. Res. Inst. 57：1—7.
11. Little, R. J. A. and D. B. Rubin. 1983. On jointly estimating parameters and missing data by maximizing the complete-da ta likelihood. Amer. Statis. 37：218—220.
12. Little, R. J. A. and D. B. Rubin. 1987. Statistical Analysis with Missing Data. John Wiley & Sons, Inc. New York. 278p.
13. Little, R. J. A. 1992. Regression with missing X's: A review. J. Amer. Statis. Assoc. 87：1227—1237.
14. Patterson, H. D. 1978. Routine least-squares estimation of variety means in incomplete tables. J. Natn. Inst. Agric. Bot. 14：401—412.
15. Perkins, J. M. and J. L. Jinks. 1968. Environmental and genotype-environmental components of variability: III. Multiple lines and crosses. Heredity 23：339—356.
16. Rubin, D. B. 1976. Comparing regression when some predictor values are missing. Technometrics 18：201—205.
17. Searle, S. R. 1987. Linear Models for Unbalanced Data. John Wiely & Sons, New York. 536p.
18. Silvery, V. 1978. Methods of analysing NIBA variety trial data over many sites and several seasons. J. Natn. Inst. Agric. Bot. 14：385—400.
19. Wang, R., Sedransk, J. and J. H. Jinn. 1992. Secondary data analysis when there are missing observations. J. Amer. Statis. Assoc. 87：952—961.
20. Wright, A. J. 1976. Bias in the estimation of regression coefficients in the analysis of genotype-environmental interaction. Heredity 37：299—303.
21. Yates, F. 1933. The analysis of replicated experiments when the field results are incomplete. Emp. J. Exp. Agric. 1：129—142.

Subset Method for Assessing Varietal Stability in Unbalanced Regional Trial Data[1]

Hsiu-Ying Lu[2]

Summary

Linear regression analysis is the most common and useful method to assess the varietal stability in a regional trial. However, a major practical limition to date, stability analysis have been a requirement of no missing data, that is, data for every genotype and environment combination or treatment have been necessary. The precison of stability analysis can be unduly influenced by removing all the data of the genotypes or environments which contain missing plot, because of the reduction in number of genotypes and environments. This study proposed a subset method for analyzing stability of unbalanced data in regional trials. Subset analyses for unbalanced data with interaction model were simple and straightforward. The set of unbalanced data was divided into two subsets of balanced data, then one used the stability analysis for these subsets. The analyses of overlapping subsets of the data revealed available information to selecting superior genotypes with common environments. This subset method for stability analysis of unbalanced regional trial data performed well when there were many common environments for genotypes.

Key words : Subset analysis, Stability, Regional trial, Unbalanced data, Regression.

---

1. Contribution No. 1869 from Taiwan Agricultural Research Institute.
2. Statistician, Department of Agronomy, TARI, Wufeng, Taichung, Taiwan, ROC.